

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re Application of:

Syuuzi KODAMA

Group Art Unit:

Serial No.:

Examiner:

Filed: March 21, 2001

For: AUTOMATIC TRANSLATOR AND COMPUTER-READABLE STORAGE  
MEDIUM HAVING AUTOMATIC TRANSLATION PROGRAM  
RECORDED THEREON



**SUBMISSION OF CERTIFIED COPY OF PRIOR  
FOREIGN APPLICATION IN ACCORDANCE WITH  
THE REQUIREMENTS OF 37 C.F.R. § 1.55**

Assistant Commissioner for Patents  
Washington, D.C. 20231

Sir:

In accordance with the provisions of 37 C.F.R. § 1.55, the applicant(s) submit(s)  
herewith a certified copy of the following foreign application(s):

Japanese Patent Application No. 2000-207714  
Filed: July 10, 2000

It is respectfully requested that the applicant(s) be given the benefit of the foreign filing  
date, as evidenced by the certified papers attached hereto, in accordance with the requirements  
of 35 U.S.C. § 119.

Respectfully submitted,  
STAAS & HALSEY LLP

Date: March 21, 2001

By: \_\_\_\_\_

  
H. J. Staas  
Registration No. 22,010

700 Eleventh Street, N.W.  
Suite 500  
Washington, D.C. 20001  
Telephone: (202) 434-1500  
Facsimile: (202) 434-1501

日 本 国 特 許 庁  
PATENT OFFICE  
JAPANESE GOVERNMENT



別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日  
Date of Application: 2000年 7月10日

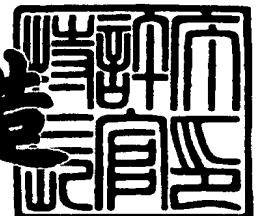
出 願 番 号  
Application Number: 特願2000-207714

出 願 人  
Applicant (s): 富士通株式会社

2000年11月17日

特許庁長官  
Commissioner,  
Patent Office

及 川 耕 造



出証番号 出証特2000-3095349

【書類名】 特許願

【整理番号】 0000020

【提出日】 平成12年 7月10日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 17/28

【発明の名称】 機械翻訳装置及び機械翻訳プログラムを記録したコンピュータ読み取り可能な記録媒体

【請求項の数】 5

【発明者】

    【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

    【氏名】 小玉 修司

【特許出願人】

    【識別番号】 000005223

    【氏名又は名称】 富士通株式会社

【代理人】

    【識別番号】 100086863

    【弁理士】

    【氏名又は名称】 佐藤 英世

【手数料の表示】

    【予納台帳番号】 061528

    【納付金額】 21,000円

【提出物件の目録】

    【物件名】 明細書 1

    【物件名】 図面 1

    【物件名】 要約書 1

    【包括委任状番号】 9906781

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 機械翻訳装置及び機械翻訳プログラムを記録したコンピュータ読み取り可能な記録媒体

【特許請求の範囲】

【請求項 1】 翻訳対象の原文を入力する原文入力手段と、

前記原文入力手段により入力された原文から形態素の列を切り出す形態素解析手段と、

前記形態素を概念範疇に変換して出力する変換手段と、

前記変換手段により出力された概念範疇の列中に所定の定型的な概念範疇の並びが存在するか否かを検証する定型文検証手段と、

前記定型文検証手段により前記概念範疇の列中に前記概念範疇の並びが存在すると判定された場合に、概念範疇の並びに対応した所定の訳文の型を生成し、該訳文の型を構成する概念範疇の原文の形態素に対応した訳語に置き換える置換手段と

を有することを特徴とする機械翻訳装置。

【請求項 2】 前記変換手段は、概念範疇とその中に含まれる形態素との関係を定義した語彙情報ファイルを備え、

前記定型文検証手段は、所定の概念範疇の並びと該並びに対応した所定の訳文の型との対を記録した第 1 のテーブルを備えると共に、

前記置換手段は、前記訳文の型を構成する形態素と該形態素に対応した所定の訳語との対を記録した第 2 のテーブルを備えたことを特徴とする請求項 1 記載の機械翻訳装置。

【請求項 3】 前記語彙情報ファイル、第 1 のテーブル、及び第 2 のテーブルの少なくとも 1 つは、ユーザが定義乃至設定ができるようにしたことを特徴とする請求項 2 記載の機械翻訳装置。

【請求項 4】 翻訳対象の原文を入力する原文入力ステップと、

前記原文入力ステップにより入力された原文から形態素の列を切り出す形態素解析ステップと、

前記形態素解析ステップで切り出された形態素を概念範疇に変換して出力する

変換ステップと、

前記変換ステップにより出力された概念範疇の列中に所定の定型的な概念範疇の並びが存在するか否かを検証する定型文検証ステップと、

前記定型文検証ステップにより前記概念範疇の列中に前記概念範疇の並びが存在すると判定された場合に、概念範疇の並びに対応した所定の訳文の型を生成し、該訳文の型を構成する概念範疇の原文の形態素に対応した訳語に置き換える置換ステップと

を実行する機械翻訳プログラムを記録したコンピュータ読み取り可能な記録媒体

。 【請求項 5】 前記変換ステップでは、概念範疇とその中に含まれる形態素との関係を定義した語彙情報ファイルを参照して、切り出された形態素を概念範疇に変換し、

前記定型文検証ステップでは、所定の概念範疇の並びと該並びに対応した所定の訳文の型との対を記録した第 1 のテーブルを参照してその検証を行い、

前記置換ステップでは、第 1 のテーブルを参照して概念範疇の並びに対応した所定の訳文の型の生成を行うと共に、前記訳文の型を構成する形態素と該形態素に対応した所定の訳語との対を記録した第 2 のテーブルを参照することでその置換を行うことを特徴とする請求項 4 記載のコンピュータ読み取り可能な記録媒体

。 【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、機械翻訳装置に関し、特に、所定の定型的な文に対し、そのような定型文に対応する翻訳文を記録したテーブルを検索する手法により、翻訳を行う機械翻訳装置に関する。併せて、そのような翻訳を行うプログラムを記録したコンピュータ読み取り可能な記録媒体についても提案するものである。

【0002】

【従来の技術】

従来の機械翻訳装置は、翻訳すべき原文全体を一様に解析処理していた。図 4

は、従来の典型的な機械翻訳装置の構成を示すブロック図である。同図に示すように、従来は、原文全体を原文入力部 3 1 で順次に入力し、これを形態素解析部 3 2 で形態素に分解した後、該形態素の列を、構文解析部 3 3、意味解析部 3 4、概念構造変換部 3 5 の順で解析処理していた。

## 【 0 0 0 3 】

すなわち、翻訳すべき原文は、形態素解析部 3 2 で形態素に分解された後、例外無く、構文解析部 3 3 以降の深いレベルの解析部を通過し、その翻訳が行われていた。

## 【 0 0 0 4 】

## 【発明が解決しようとする課題】

原文が定型的な文であり、その翻訳結果が原文との単純な対応関係だけで決定できるようなケースに対しては、上記した従来の翻訳処理では、明らかにオーバースペックであった。

## 【 0 0 0 5 】

すなわち、深いレベルの解析部を通す原文の長さが少しも短縮されないので、原文の長さに略比例した誤り翻訳の可能性を残すことになり、又、翻訳に要する処理時間も短縮されないことになる。

## 【 0 0 0 6 】

特に、近年のインターネットの爆発的普及と共に、Eメールやホームページの閲覧の際に、機械翻訳装置が多用されることとなったが、このような場合には、翻訳結果が単純な対応関係だけで決定できるような原文が、所与の翻訳対象に含まれているケースが多くなっている。その代表例は、"Thu, 09 Jul 1998 07:49:30 -0700"のような日時や曜日を表す文字列である。

## 【 0 0 0 7 】

このような文字列には、複数通りのパターンが存在し得るが、それらは典型的であり、強い規則性を持っている。より具体的に言えば、このような文字列は、本質的には名詞相当の語の単なる羅列であり、通常の文のように句や節といった構造を持つものではない。

## 【 0 0 0 8 】

従って、上記の文字列を入力した時に、図 3 に示す構文解析部 3 3、意味解析部 3 4、概念構造変換部 3 5 の各処理を順次を実施することは、明らかにオーバーサックであり、このため、誤解析の可能性が常に残され、かつ、翻訳速度の点からも望ましいものではない。

【 0 0 0 9 】

一方、特開昭 6 4 - 3 1 2 6 9 号公報では、オペレータにより指定された単語およびその訳語を記憶する訳語指定メモリを有し、翻訳実行時に前記訳語指定メモリを参照することにより、原文中の前記訳語指定メモリに記憶されている単語についてはそれに対応した指定の訳語を用いて訳文を生成することを開示している。

【 0 0 1 0 】

しかし、同公報に開示された翻訳装置は、単語から単語への単純な対応による翻訳処理であり、形態素の列中から所定の形態素の並び（一般には原文の複数の単語に対応）を検出などの処理をしておらず、オペレータによる事前の訳語登録処理を必要とするなど、ユーザインターフェースは極めて稚拙なものであった。

【 0 0 1 1 】

本発明は、以上のような従来の機械翻訳装置における問題点に鑑みてなされたものであり、定型的な文に対して、パターンマッチング手法により高速に解析処理することができる機械翻訳装置及びそのような翻訳を行うプログラムを記録したコンピュータ読み取り可能な記録媒体を提供せんとするものである。

【 0 0 1 2 】

【課題を解決するための手段】

上記課題を解決するために、本発明に係る機械翻訳装置は、  
 翻訳対象の原文を入力する原文入力手段と、  
 前記原文入力手段により入力された原文から形態素の列を切り出す形態素解析手段と、  
 前記形態素を概念範疇に変換して出力する変換手段と、  
 前記変換手段により出力された概念範疇の列中に所定の定型的な概念範疇の並びが存在するか否かを検証する定型文検証手段と、

前記定型文検証手段により前記概念範疇の列中に前記概念範疇の並びが存在すると判定された場合に、概念範疇の並びに対応した所定の訳文の型を生成し、該訳文の型を構成する概念範疇の原文の形態素に対応した訳語に置き換える置換手段と

を有することを基本的特徴としている（請求項1）。

#### 【0013】

以上の構成では、原文入力手段から、例えば、“Thu, 09 Jul 1998 07:49:30 - 0700”のような日時や曜日を表す文字列が入力されると、形態素解析手段により、形態素解析を行って、「Thu」「,」「09」「Jul」「1998」「07:49:30」「-0700」という形態素に変換される。

#### 【0014】

このままでは、本発明の構成による翻訳（置換処理）を行って良いか否かが明らかではない。元来形態素列というのは、外形的には元の文字列を最小限の意味単位で区切った表記の連なりであると同時に、意味的には概念範疇の連なりであるとみなすことができる。ここで、概念範疇とは、形態素をその概念によって分類する分類先の所定のカテゴリー、すなわち意味情報の一種を言う。

#### 【0015】

上記変換手段による形態素から概念範疇への変換は、切り出された形態素を、特定のライブラリなどで構成される語彙情報ファイルに照らし合わせて、概念範疇として取り出すことになる。以上の語彙情報ファイルには、所定の概念範疇に対する定義が行われており（例えば全ての要素を列挙する外縁的定義や、各要素が満たすべき制約を記述する内包的定義など）、対象となる形態素がこれらの定義の条件を満たす場合に、該定義で規定される概念範疇が、上記ファイルから抽出される。

#### 【0016】

従ってこの語彙情報ファイルにおける概念範疇をユーザが自由に定義できるユーザインターフェースを設けておけば、色々な分野の定型的な文に対しても本構成による自動翻訳が可能となる。

#### 【0017】



本構成では、上記変換手段を使って、「Thu」「,」「09」「Jul」「1998」「07:49:30」「-0700」と切り出された形態素を、意味情報を示す概念範疇に変換し（例えば「Thu」→〔省略された曜日名〕）、その概念範疇列が、一定の規則性を持つか否か、すなわち上記概念範疇の列中に所定の定型的な概念範疇の並びが存在するか否かを、前記定型文検証手段によって、検証する。上記例において、いくつか用意されたパターンの中から、〔省略された曜日名〕〔カンマ〕〔日〕〔省略された月名〕〔4桁でかかれた年〕……という特定のパターンに合致することが定型文検証手段により検証されたものとすれば、前記置換手段により、概念範疇の並びに対応した所定の訳文の型を生成し（上記の例で言えば、〔4桁でかかれた年〕年〔省略された月名〕月〔日〕日（〔省略された曜日名〕）という型）、該訳文の型を構成する概念範疇の原文の形態素（上記の例で言えば、「1998」「Jul」「09」「Thu」）に対応した訳語に置き換える（上記の例で言えば、1998年07月09日（木）という訳語）置換処理が行われることになる。

## 【 0 0 1 8 】

また、前記定型文検証手段は、所定の概念範疇の並びと該並びに対応した所定の訳文の型との対を記録した第1のテーブルを備え、このテーブルと前記概念範疇列とのパターンマッチングにより、定型的な概念範疇の並びが存在するか否かを検証する。

## 【 0 0 1 9 】

他方前記置換手段は、前記訳文の型を構成する形態素と該形態素に対応した所定の訳語との対を記録した第2のテーブルを備え、前記第1のテーブルを使用して、概念範疇の並びに対応した所定の訳文の型を生成すると共に、第2のテーブルを使用して、該訳文の型を構成する概念範疇の原文の形態素に対応した訳語に置き換える。

## 【 0 0 2 0 】

前記語彙情報ファイルだけでなく、第1のテーブルでは、所定の概念範疇の並びと該並びに対応した所定の訳文の型との対を、また第2のテーブルでは、訳文の型を構成する形態素と該形態素に対応した所定の訳語との対を、ユーザが自由に設定ができるようにすると、色々な分野の定型的な文に対しても本構成による

自動翻訳が可能となる。

#### 【 0 0 2 1 】

上記置換手段による訳文の型の生成において、該型には、必要に応じて形態素の並びを所定の順序に変え、且つその時必要な添え字を付すことも可能である。これは、例えば、“Thu, 09 Jul 1998”のような日時や曜日を表す文字列があった場合、「(木), 09 07 1998」と単に置換するだけでなく、その並びを変え且つ添え字を付けて、「1998年07月09日(木)」と語順を変える方が、日本語的には分かり易くなるからである。従って、訳文の型の生成時は、上記例では、〔4桁でかかれた年〕年〔省略された月名〕月〔日〕日（〔省略された曜日名〕）といった型が生成されることになる。これをに基に、該訳文の型を構成する概念範疇の原文の形態素に対応した訳語に置き換えることで、「1998年07月09日(木)」という翻訳が行われることになる。

#### 【 0 0 2 2 】

さらに請求項4及び請求項5の構成は、請求項1及び請求項2に記載の構成を、コンピュータなどの計算装置に実行させるために、それらの装置で実行可能なプログラムを格納した記録媒体を規定している。すなわち、請求項4の構成では、

翻訳対象の原文を入力する原文入力ステップと、

前記原文入力ステップにより入力された原文から形態素の列を切り出す形態素解析ステップと、

前記形態素解析ステップで切り出された形態素を概念範疇に変換して出力する変換ステップと、

前記変換ステップにより出力された概念範疇の列中に所定の定型的な概念範疇の並びが存在するか否かを検証する定型文検証ステップと、

前記定型文検証ステップにより前記概念範疇の列中に前記概念範疇の並びが存在すると判定された場合に、概念範疇の並びに対応した所定の訳文の型を生成し、該訳文の型を構成する概念範疇の原文の形態素に対応した訳語に置き換える置換ステップと

を実行するプログラムを記録したコンピュータ読み取り可能な記録媒体を開示す

る。

【 0 0 2 3 】

また請求項 5 の構成において、

前記変換ステップでは、概念範疇とその中に含まれる形態素との関係を定義した語彙情報ファイルを参照して、切り出された形態素を概念範疇に変換し、

前記定型文検証ステップでは、所定の概念範疇の並びと該並びに対応した所定の訳文の型との対を記録した第 1 のテーブルを参照してその検証を行い、

前記置換ステップでは、第 1 のテーブルを参照して概念範疇の並びに対応した所定の訳文の型の生成を行うと共に、前記訳文の型を構成する形態素と該形態素に対応した所定の訳語との対を記録した第 2 のテーブルを参照することでその置換を行うことを開示している。

【 0 0 2 4 】

この場合、上記計算装置とは、中央演算処理装置の構成を含んだ汎用的なコンピュータの構成の他、特定の処理に向けられた専用機（自動翻訳機など）などであっても良く、中央演算処理装置の構成を伴うものであれば、特に限定はない。

【 0 0 2 5 】

このような記録媒体から、上記プログラムが計算装置に読み出され、実行されると、上記各ステップが実行されることになる。

【 0 0 2 6 】

さらに、前記語彙情報ファイル、第 1 のテーブル、及び第 2 のテーブルの少なくとも 1 つは、ユーザが定義乃至設定ができるようにすると良い。そうすることで、前述したように、色々な分野の定型的な文に対しても本構成による自動翻訳が可能となる。

【 0 0 2 7 】

尚、請求項 4 及び請求項 5 記載の各手順のうちその一部は、計算装置に組み込まれた手順（計算装置にハードウェア的に組み込まれている手順でも良く、該計算装置に組み込まれているオペレーティングシステムや他のアプリケーションプログラムなどによって実現される手順でも良い）によって実現され、前記記録媒体に記録されたプログラムには、該計算装置によって実現される手順を呼び出す

あるいはリンクさせる命令が含まれていても良い。

【 0 0 2 8 】

これは、請求項 4 及び請求項 5 に規定された各手順の一部が、例えばオペレーティングシステムなどの手順の一部で代行され、記録媒体にはその手順を実現するためのプログラムないしモジュールなどは直接記録されているわけではないが、それらの手順を達成するオペレーティングシステム手順の一部を、呼び出したりリンクさせるようにしてあれば、実質的に同じ構成となるからである。

【 0 0 2 9 】

【発明の実施の形態】

以下、本発明の実施の形態を図面を参照して説明する。なお、以後の説明では、記録媒体である CD-ROM に、機械翻訳を実行するためのプログラムが記録されており、この CD-ROM をパソコンやその他の汎用機に読み込ませることで、これらの装置が、本発明の機械翻訳装置として提供される。ただしこのような本実施形態の構成に限られず、自動翻訳専用機として構成することも可能である。

【 0 0 3 0 】

図 1 は、本発明の実施の形態に係る機械翻訳装置のシステム構成を示すシステム概要図である。同図に示すように、本システム 1 0 0 は、CPU 1 0 2、RAM 1 0 4、ROM 1 0 6、HDD 1 0 8 などの外部記憶装置、同じく外部記録装置の一部を構成する CD 読み込み部 1 1 0、画像処理部 1 1 2 a を介してつながる CRT 1 1 2、入力装置 1 1 4、その他の外部出力装置 1 1 6、及び通信デバイス部 1 1 8 により構成されており、各部はバスによって接続されている。

【 0 0 3 1 】

上記 CPU 1 0 2 は、CD 読み込み部 1 1 0 を介して、CD-ROM から読み出されるプログラムや、HDD 1 0 8 などに格納されているデータに基づいて、プログラムの実行や本システム 1 0 0 の各部の制御など種々な処理を行うものである。

【 0 0 3 2 】

上記 RAM 1 0 4 は、CD 読み込み部 1 1 0 を介して CD-ROM から読み出し

たプログラムを含む各種データや、CPU 1 0 2 の処理結果などを記憶する格納領域を形成する。

【 0 0 3 3 】

上記HDD 1 0 8 は、ユーザのインストール作業により、CD-ROMに格納されたプログラム及びデータを、CD読み込み部 1 1 0 から読み出し、格納しておく構成であり、また必要に応じて、CPU 1 0 2 の作業領域を提供する。さらにHDD 1 0 8 自身が、前記インストールにより、本発明の請求項 3 及び請求項 4 の記録媒体ともなることがある。

【 0 0 3 4 】

CD読み込み部 1 1 0 は、CD-ROMからプログラム及びデータの読み出しを行うものであり、ハードウェア的には、CD-ROMドライブとCD-ROMバッファにより構成される。CD-ROMには、本発明を実現するためのプログラムを含む機械翻訳プログラムおよび各種データ（各種辞書や後述するテーブルデータを含む）が格納されている。

【 0 0 3 5 】

画像処理部 1 1 2 a は、CPU 1 0 2 での処理結果に基づいて、文字キャラクタやその他の画像データを生成する処理を行うものである。画像処理部 1 1 2 a で生成された画像データはCRT 1 1 2 に出力され、表示される。

【 0 0 3 6 】

入力装置 1 1 4 は、キーボード及びポインティングデバイス（マウスなど）で構成されており、ユーザが必要に応じて、文字の入力やデータの入力、或いは各種の指示の入力（コマンドの入力を含む）を行う。

【 0 0 3 7 】

その他の外部出力装置 1 1 6 は、プリンタで構成され、ユーザが必要に応じて、CRT 1 1 2 で出力される処理結果などのプリントアウトを行う。

【 0 0 3 8 】

通信デバイス部 1 1 8 は、本システム 1 0 0 と、他のシステム構成（図示なし）との間でローカルエリアネットワーク（LAN）を実現するための構成であり、該構成によって、プログラム、各種コマンド或いは各種のデータを転送する。

## 【 0 0 3 9 】

本構成では、本システム 1 0 0 の電源スイッチを ON にして、CD-ROM をセットすると、上記本システム 1 0 0 の各部の初期化が行われ、プログラムロードレディとなったところで、CPU 1 0 2 により CD 読み込み部 1 1 0 から本発明構成を含む機械翻訳プログラム及び関連データが読み出される。そしてこれらは、一旦 HDD 1 0 8 にインストールされる。さらに該 HDD 1 0 8 から、前記 CPU 1 0 2 に読み出されて RAM 1 0 4 に格納されたプログラムは、CPU 1 0 2 で実行されることで、本発明の機械翻訳装置が実現される。

## 【 0 0 4 0 】

図 2 は、CPU 1 0 2 により記録媒体である前記 CD-ROM から本発明の構成を実現するための翻訳プログラムが読み出され実行された場合に、上記本システム 1 0 0 上に生成される本実施形態の機械翻訳装置の全体構成を示すブロック図である。

## 【 0 0 4 1 】

すなわち本実施形態に係る機械翻訳装置は、本翻訳プログラムと一緒に本システム 1 0 0 上で立ち上がる別の翻訳ソフトからの原文を入力する原文入力部 1 と、原文入力部 1 に入力された原文を形態素に切り出す形態素解析部 2 と、形態素解析部 2 から切り出された形態素を概念範疇に変換する変換部 3 と、変換部 3 により変換された概念範疇の列中に所定の定型的な概念範疇の並びが存在するか否かを検証する定型文検証部 4 と、上記概念範疇の列中に定型的な概念範疇の並びが存在する時に、概念範疇の並びに対応した所定の訳文の型を生成し、該訳文の型を構成する概念範疇の原文の形態素に対応した訳語に置き換える置換部 5 とを有している。

## 【 0 0 4 2 】

さらに前記変換部 3 には、それによって参照される、概念範疇とその中に含まれる形態素との関係を定義した語彙情報ファイル 6 が備えられ、また前記定型文検証部 4 には、同じくそれによって参照される、所定の概念範疇の並びと該並びに対応した所定の訳文の型との対を記録したパターンマッチングテーブル 7 が備えられ、さらに前記置換部 5 には、同じくそれによって参照される、前記訳文の

型を構成する形態素と該形態素に対応した所定の訳語との対を記録した置き換えテーブル 8 が備えられている。

#### 【 0 0 4 3 】

以下、本実施の形態に係る機械翻訳装置の基本的な機能を説明する。

翻訳対象の原文は、前述した別の翻訳ソフトから、カット＆ペーストなどの操作をユーザが行うことにより、クリップボードにコピーされ、該クリップボードを介して原文入力部 1 により、本システム 1 0 0 の機械翻訳装置に入力される。上記入力処理は、原文全体を一括入力してもよいし、論理的又は物理的な観点により、適当な入力単位に分割して入力してもよい。

#### 【 0 0 4 4 】

上記入力された原文は、形態素解析部 2 により形態素の列に分解（切り出し）され、上記切り出された形態素の列は、前記変換部 3 に入力される。

#### 【 0 0 4 5 】

該変換部 3 では、前記形態素を概念範疇に変換して出力する。前述のように、形態素列に分解されただけでは、本システム 1 0 0 構成による翻訳（置換処理）を行って良いか否かが明らかではない。その点、形態素列というのは、外形的には元の文字列を最小限の意味単位で区切った表記の連なりであると同時に、意味的には概念範疇の連なりであるとみなすことができる。そして後述する定型文検証部 4 により、この概念範疇の並びを見、本システム 1 0 0 で翻訳を行って良いか否かの判断を行う前提として、本変換部 3 で、形態素を概念範疇に変換し、概念範疇の列を作成する。

#### 【 0 0 4 6 】

この変換部 3 による形態素から概念範疇への変換は、切り出された形態素を、所定のライブラリで構成されるにある上記語彙情報ファイル 6 に照らし合わせて、概念範疇として取り出すことになる。該語彙情報ファイル 6 には、所定の概念範疇に対する定義が行われており（例えば全ての要素を列挙する外縁的定義や、各要素が満たすべき制約を記述する内包的定義など）、対象となる形態素がこれらの定義の条件を満たす場合に、該定義で規定される概念範疇が抽出される。本構成では、この語彙情報ファイル 6 における概念範疇をユーザが自由に定義でき

るユーザインターフェースが設けられており、形態素から概念範疇への変換率を向上せしめている。

## 【 0 0 4 7 】

前記定型文検証部 4 では、意味情報を示す概念範疇の列が、一定の規則性を持つか否か、すなわち上記概念範疇の列中に所定の定型的な概念範疇の並びが存在するか否かを、検証する。この検証処理に際しては、定型文検証部 4 は、概念範疇の列、又はその任意の部分列をキーとして、上記のパターンマッチングテーブル 7 を検索する。

## 【 0 0 4 8 】

上記検索処理を含む検証の結果、上記概念範疇の列中に、所定の定型的な概念範疇の並びが存在しない場合には、該概念範疇の列に対応する元の形態素の列は、全て前述した別の翻訳ソフトにおける深いレベルの解析部に回される。尚、この深いレベルの解析部は、前述した従来技術の機械翻訳装置の構成と同じであるので、説明を省略する。

## 【 0 0 4 9 】

上記検証の結果、上記概念範疇の列中に、所定の定型的な概念範疇の並びが存在する場合には、この概念範疇の列は、置換部 5 に送られ、該置換部 5 において、概念範疇の並びに対応した所定の訳文の型が生成される。前記パターンマッチングテーブル 7 には、定型的な概念範疇の並び（パターン）のみが記録されているわけではなく、その並びに対応した所定の訳文の型が対になって記録されている。従って、前記置換部 5 は、パターンマッチングテーブル 7 を参照して概念範疇の並びに対応した所定の訳文の型の生成を行う。

## 【 0 0 5 0 】

さらに置換部 5 は、訳文の型を構成する形態素と該形態素に対応した所定の訳語との対を記録した置き換えテーブル 8 を参照して、前記訳文の型を構成する概念範疇の原文の形態素に対応した訳語に置き換える。

## 【 0 0 5 1 】

他方、上記概念範疇の列中の上記所定の定型的な概念範疇の並びが存在する部分を除く部分は、当該定型的な概念範疇の並びが存在する箇所が認識できるよう



に識別コードを挿入する等の細工を施してから、全て上記の深いレベルの解析部に回される。

## 【 0 0 5 2 】

図 3 は、本発明の実施の形態に係る機械翻訳装置の具体例による動作を示すフローチャートである。以下、図 2 を参照しつつ、図 3 のフローチャートを使用して、本実施の形態に係る機械翻訳装置の具体例による動作を説明する。

## 【 0 0 5 3 】

まず、ステップ S 1 では、原文入力部 1 が、翻訳対象となる原文を入力する。

## 【 0 0 5 4 】

次に、ステップ S 2 では、形態素解析部 2 が、上記入力された原文の全体又は所定の処理単位を解析して、これを形態素の列に分解（切り出し）した後、変換部 3 に送出する。そしてステップ S 3 では、該変換部 3 が、語彙情報ファイル 6 を参照して、個々の形態素を概念範疇に変換する。

## 【 0 0 5 5 】

さらに、ステップ S 4 では、定型文検証部 4 が、パターンマッチングテーブル 7 を検索して、変換された概念範疇の列中に、定型的な概念範疇の並びが存在するか否かを検証する。具体的には、概念範疇の列、又はその任意の部分列をキーとしてパターンマッチングテーブル 7 を検索する。

## 【 0 0 5 6 】

ここで、概念範疇の列中に、上記の定型的な概念範疇の並びが存在しなければ、ステップ S 6 で、分解された形態素の列を深いレベルの解析処理に回す。

## 【 0 0 5 7 】

他方、上記の定型的な概念範疇の並びが存在する時には、概念範疇の列を置換部 5 に送る。ステップ S 5 では、該置換部 5 において、前記パターンマッチングテーブル 7 を参照し、該テーブル 6 に定型的な概念範疇の並び（パターン）と対になって記録されている、該並びに対応した所定の訳文の型として取り出すことで、所定の訳文の型の生成を行う。そして前記置き換えテーブル 8 を参照して、訳文の型を構成する概念範疇の原文の形態素に対応した訳語に置き換える。

## 【 0 0 5 8 】

上記の置換処理が完了した翻訳文は、最終的な訳文生成を実施する別の翻訳ソフトの実行処理部に、前記クリップボードなどを介して送られるが、該翻訳ソフトについては、従来の機械翻訳装置に採用されている解析部と同じであるので、説明を省略する。

## 【 0 0 5 9 】

以上のように、翻訳の対象となる原文のうち、上記日時などを表す定型的な部分は、本構成により、翻訳が行われる。また前記定型文検証部 4 により本構成による翻訳の対象外となると判断された部分は、本構成と一緒に立ち上げられた別の翻訳ソフトの実行処理部においてなされることになる（ステップ S 6）。そして、その両方の翻訳結果を足し合わせて訳文生成を行い、翻訳文として出力することになる（ステップ S 7）。

## 【 0 0 6 0 】

以下、上記した語彙情報ファイル 6 により形態素から概念範疇への変換を行う変換部 3 と、パターンマッチングテーブル 7 を参照する定型文検証部 4 及び置換部 5 の動作を具体例を用いて説明する。

## 【 0 0 6 1 】

この具体例では、上記変換部 3 において、上記形態素が上記の概念範疇に変換された後、さらに定型文検証部 4 において、該概念範疇の列中において、所定の定型的な概念範疇の並びの存在が検証されるケースを説明する。

## 【 0 0 6 2 】

下記の表 1 は、上記の翻訳すべき定型文が日付や曜日に関するものである場合の、上記した概念範疇の語彙情報を記録した語彙情報ファイル 6 の一覧表である。

## 【 0 0 6 3 】

【表 1】

名称	説 明	属性	要 素
曜長	省略されない曜日名	文字列	Sunday,Monday,Tuesday…
曜短	省略された曜日名	文字列	Sun,Mon,Tue…
月長	省略されない月名	文字列	January,February,March…
月短	省略された月名	文字列	Jan,Feb,Mar…
日02	日	数値	2桁の数値(範囲:1～31)
年長	4桁で書かれた年	数値	4桁の数値
年短	下2桁で書かれた年	数値	2桁の数値

## 【0064】

変換部3において、形態素を、一旦、このような概念範疇に変換させるのは、以下の理由による。すなわち、原文の形態素解析により抽出された形態素のままでは、本発明の構成による簡略化された翻訳処理を行って良いか否かが明らかではなく、形態素よりももっと意味的に抽象化されたもの（上位の概念）にして、それが一定の並びを有するか否かを定型文検証部4で判断させるために必要であるからである。

## 【0065】

形態素列というのは、元々、外形的には元の文字列を最小限の意味単位で区切った表記の連なりであると同時に、意味的には概念範疇の連なりであるとみなすことができる。従って上記変換部3による形態素から概念範疇への変換は、切り出された形態素を、特定のライブラリで構成される語彙情報ファイル6に照らし合わせて、概念範疇として取り出すことで実施される。上述のように、語彙情報ファイル6には、所定の概念範疇に対する定義が行われており（例えば全ての要素を列挙する外縁的定義や、各要素が満たすべき制約を記述する内包的定義など）、対象となる形態素がこれらの定義の条件を満たす場合に、該定義で規定される概念範疇が抽出される。

## 【0066】

変換部 3 は、上記したパターンマッチングテーブル 7 を参照する定型文検証部 4 による定型文の検証のための準備段階として、上記の表 1 の語彙情報ファイル 6 を参照しながら上記の形態素を、概念範疇に変換する。こうして変換された形態素の列は、概念範疇の列として取り出される。

## 【0067】

上記の変換処理は、上記したとおり、変換可能な形態素が連続して出現する限りにおいて継続されるので、変換不可能な形態素が出現した時点で、概念範疇の列として取り出される。このような概念範疇の列の切り出しは、上記の形態素の列の全てを対象として実施される。従って、上記切り出される概念範疇の列が複数となることもある。

## 【0068】

更に、定型文検証部 4 は、上記切り出された 1 つ又は 2 つ以上の概念範疇の列の各々をそれぞれキーとして、下記の表 2 のパターンマッチングテーブル 7 に予め記録した所定の概念範疇の並びの各々と比較し、マッチングするものが存在するか否かを検証する。

## 【0069】

【表 2】

英	〔曜短〕〔カンマ〕〔日02〕〔月短〕〔年長〕
日	〔年長〕年〔月短〕月〔日02〕日〔〔曜短〕〕
英	〔月短〕〔日02〕〔カンマ〕〔年長〕
日	〔年長〕年〔月長〕月〔日02〕日
	・ ・ ・ ・ ・

## 【0070】

上記の検証処理において、上記切り出された 1 つ又は 2 つ以上の概念範疇の並びのいずれかとマッチングする所定の概念範疇の並びが存在すれば、その結果を

前記置換部 5 に渡す。

【0071】

該置換部 5 は、上記パターンマッチングテーブル 7 を更に参照し、当該所定の概念範疇の並び（表 2 で、英として記録されている概念範疇の並び）に対応する訳文の型として、表 2 に予め記録されている対応する型（表 2 で、日として記録されている訳文）を出力する。

【0072】

この訳文の型には、必要に応じて形態素の並びを所定の順序に変え、且つその時必要な添え字が付されている。例えば、表 2 の最初の段の英の部分にある〔曜短〕〔カンマ〕〔日 0 2〕〔月短〕〔年長〕という概念範疇の並びは、その下の日の部分にある〔年長〕〔月短〕〔日 0 2〕〔曜短〕の並び順に直し、且つ各概念範疇の後部乃至周囲に、「年」「月」「日」「（ ）」などの添え字を付ける。

【0073】

そして置換部 5 は、この訳文の型を構成する概念範疇の原文の形態素に対応した訳語に置き換える。すなわち、この訳文の型を構成する概念範疇（表 2 の上段の例では〔年長〕〔月短〕〔日 0 2〕〔曜短〕）では、抽象的な概念のままであり、そのまま出力することはできない。そこで、これらの概念範疇を原文の形態素に戻し、各形態素を、それに対応した訳語に置換する。上記置換処理に際しては、下記表 3 及び表 4 で例示する置き換えテーブル 8 が使用される。

【0074】

【表 3】

〔曜短〕

日	英
日 月 火	Sun Mon Tue

【0075】

【表4】

〔曜長〕

日	英
日曜日	Sunday
月曜日	Monday
火曜日	Tuesday

【0076】

例えば、“Thu, 09 Jul 1998”のような日時や曜日を表す文字列があった場合、変換部3によって変換された概念範疇の列としては、〔曜短〕〔カンマ〕〔日02〕〔月短〕〔年長〕となり、これは、定型文検証部4によって前記パターンマッチングテーブル7に照らし合わせると、上段の英の部分にマッチングする。従って、置換部5により、表2上段の英の直ぐ下にある日の部分〔年長〕年〔月短〕月〔日02〕日（〔曜短〕）を、訳文の型として生成する。さらにこれらの概念範疇を原文の形態素に戻すと、「1998」「Jul」「09」「Thu」となる。これらの形態素のうち、“Jul”や“Thu”は、英語表記の短縮形であり、翻訳がされていない。これらを、前記表3及び表4などの置き換えテーブル8を参照して、夫々、“Jul”→“7”、“Thu”→“木”に置換する。その結果、1998年07月09日（木）として出力されることになる。

【0077】

上記の実施の形態において、深いレベルの解析部と訳文生成部については、本発明の構成とは別に用意され、一緒に立ち上げられた翻訳ソフトにおいてなされることになる。すなわち、翻訳の対象となる原文のうち、上記日時などを表す定型的な部分は、本構成により、翻訳が行われ、また前記定型文検証部4により本構成による翻訳の対象外となると判断された部分は、前記一緒に立ち上げられた翻訳ソフトにおいてなされることになる。そして、その両方の翻訳結果を足し合わせて訳文生成を行い、翻訳文として出力することになる。

## 【 0 0 7 8 】

上記深いレベルの解析部と訳文生成部については、本発明の構成とは別の構成（ソフト）においてなされると説明したが、上記の深いレベルの解析部と訳文生成部とを本発明の構成要素に含めることも可能である。

## 【 0 0 7 9 】

また、定型文検証部 4 及び置換部 5 が参照する前記の各表は、表の先頭エントリからの順次検索に対応した構成とは限らず、（例えば、格納すべきエントリ内容をアルファベット順で並べたり、適当なハッシュ関数を作用させたりして、）検索時に高速検索が可能となるような工夫を施してもよい。

## 【 0 0 8 0 】

さらに、図 3 のフローチャートで示した各処理を行うプログラムは、CD-ROM に格納されて提供される場合だけでなく、フロッピーディスク、磁気テープなどの、コンピュータ読み取り可能な記録媒体に格納して配付しても良い。そして、少なくとも、パーソナルコンピュータ、汎用コンピュータをその範疇に含むコンピュータが、上記の記録媒体から上記プログラムを読み出して、実行するものとしても良い。

## 【 0 0 8 1 】

尚、本発明の機械翻訳装置は、上述の実施例にのみ限定されるものではなく、本発明の要旨を逸脱しない範囲内において種々変更を加え得ることは勿論である。

## 【 0 0 8 2 】

たとえば、上記定型文検証部 4 において備えられたパターンマッチングテーブル 7 には、所定の概念範疇の並びのみが記録されており、該並びに対応した所定の訳文の型は、前記置換部 5 の別のテーブルとして備えられ、定型文検証部 4 において、パターンマッチングテーブル 7 におけるマッチングで、定型的な概念範疇の並びの存在が検証された場合に、置換部 5 は、該パターンマッチングテーブル 7 とは別に用意された前記テーブルを参照して、概念範疇の並びに対応した所定の訳文の型の生成を行うことも可能である。

## 【 0 0 8 3 】

さらに上記定型文検証部 4 と置換部 5 との構成を 1 つの構成とすることも可能である。1 つの構成とした場合、定型的な概念範疇の並びの存在の検証と、概念範疇の並びに対応した所定の訳文の型の生成、及び訳文の型を構成する概念範疇の原文の形態素に対応した訳語への置き換えなどの処理が一緒になされることになる。その方が、定型的な概念範疇の並びの存在の検証と、概念範疇の並びに対応した所定の訳文の型の生成を行う場合に参照されるテーブルが、前記したパターンマッチングテーブル 7 のような構成で済み、別々に分ける必要がなくなって、効率的である。

【 0 0 8 4 】

(付記 1) 翻訳対象の原文を入力する原文入力手段と、  
前記原文入力手段により入力された原文から形態素の列を切り出す形態素解析手段と、  
前記形態素を概念範疇に変換して出力する変換手段と、  
前記変換手段により出力された概念範疇の列中に所定の定型的な概念範疇の並びが存在するか否かを検証する定型文検証手段と、  
前記定型文検証手段により前記概念範疇の列中に前記概念範疇の並びが存在すると判定された場合に、概念範疇の並びに対応した所定の訳文の型を生成し、該訳文の型を構成する概念範疇の原文の形態素に対応した訳語に置き換える置換手段と  
を有することを特徴とする機械翻訳装置。

(付記 2) 前記変換手段は、概念範疇とその中に含まれる形態素との関係を定義した語彙情報ファイルを備え、

前記定型文検証手段は、所定の概念範疇の並びと該並びに対応した所定の訳文の型との対を記録した第 1 のテーブルを備えると共に、

前記置換手段は、前記訳文の型を構成する形態素と該形態素に対応した所定の訳語との対を記録した第 2 のテーブルを備えたことを特徴とする付記 1 記載の機械翻訳装置。

(付記 3) 前記語彙情報ファイル、第 1 のテーブル、及び第 2 のテーブルの少なくとも 1 つは、ユーザが定義乃至設定ができるようにしたことを特徴とする



付記 2 記載の機械翻訳装置。

(付記 4) 翻訳対象の原文を入力する原文入力ステップと、

前記原文入力ステップにより入力された原文から形態素の列を切り出す形態素解析ステップと、

前記形態素解析ステップで切り出された形態素を概念範疇に変換して出力する変換ステップと、

前記変換ステップにより出力された概念範疇の列中に所定の定型的な概念範疇の並びが存在するか否かを検証する定型文検証ステップと、

前記定型文検証ステップにより前記概念範疇の列中に前記概念範疇の並びが存在すると判定された場合に、概念範疇の並びに対応した所定の訳文の型を生成し、該訳文の型を構成する概念範疇の原文の形態素に対応した訳語に置き換える置換ステップと

を実行する機械翻訳プログラムを記録したコンピュータ読み取り可能な記録媒体。

(付記 5) 前記変換ステップでは、概念範疇とその中に含まれる形態素との関係を定義した語彙情報ファイルを参照して、切り出された形態素を概念範疇に変換し、

前記定型文検証ステップでは、所定の概念範疇の並びと該並びに対応した所定の訳文の型との対を記録した第 1 のテーブルを参照してその検証を行い、

前記置換ステップでは、第 1 のテーブルを参照して概念範疇の並びに対応した所定の訳文の型の生成を行うと共に、前記訳文の型を構成する形態素と該形態素に対応した所定の訳語との対を記録した第 2 のテーブルを参照することでその置換を行うことを特徴とする付記 4 記載のコンピュータ読み取り可能な記録媒体。

(付記 6) 前記語彙情報ファイル、第 1 のテーブル、及び第 2 のテーブルの少なくとも 1 つは、ユーザが定義乃至設定ができるようにしたことを特徴とする付記 5 記載のコンピュータ読み取り可能な記録媒体。

【 0 0 8 5 】

【発明の効果】

以上説明した本発明によれば、原文に定型的な文が含まれ、該定型的な文の翻

訳結果が、該定型的な文とその訳文との単純な対応関係だけで決定できるようなケースに対しては、その部分だけ本構成により翻訳が完了し、上記深いレベルの解析部に回す必要がなくなるため、原文全体の翻訳において、深いレベルの解析部に回す原文の長さを短縮することが可能になる。その結果誤訳の発生する可能性が減少し、かつ、翻訳に要する処理時間も短縮することができる。

## 【 0 0 8 6 】

また特開昭 6 4 - 3 1 2 6 9 号公報に開示された翻訳装置のように、単語から単語への単純な対応による翻訳処理を行っているわけではなく、形態素を一旦概念範疇という上位の概念に直して、この概念範疇の列中から所定の概念範疇の並びを検出するなどの処理を行っており、それによって本構成による高速且つ誤りの少ない翻訳を自動的にすることが可能となるなど、オペレータによる事前の訳語登録処理などをしなくて済み、ユーザインターフェースも格段に向上する。

## 【 0 0 8 7 】

さらに形態素を概念範疇に変換するための語彙情報ファイル、パターンマッチングテーブル、置き換えテーブルなどの仕様を明らかにして、ユーザ側で自由に定義乃至設定できるようにすれば、本実施例で示した日時や日付を表す定型文の場合だけでなく、その他の（その他の分野の）定型的な文に対しても本構成による自動翻訳が可能となり、誤訳の可能性の減少及び翻訳スピードの点で、さらに有利となる。

## 【図面の簡単な説明】

## 【図 1】

本発明の実施の形態に係る機械翻訳装置のシステム構成を示すシステム概要図である。

## 【図 2】

本発明の実施の形態に係る機械翻訳装置の全体構成を示すブロック図である。

## 【図 3】

本発明の実施の形態に係る機械翻訳装置の具体例による動作を示すフローチャートである。

## 【図 4】

従来の典型的な機械翻訳装置の構成を示すブロック図である。

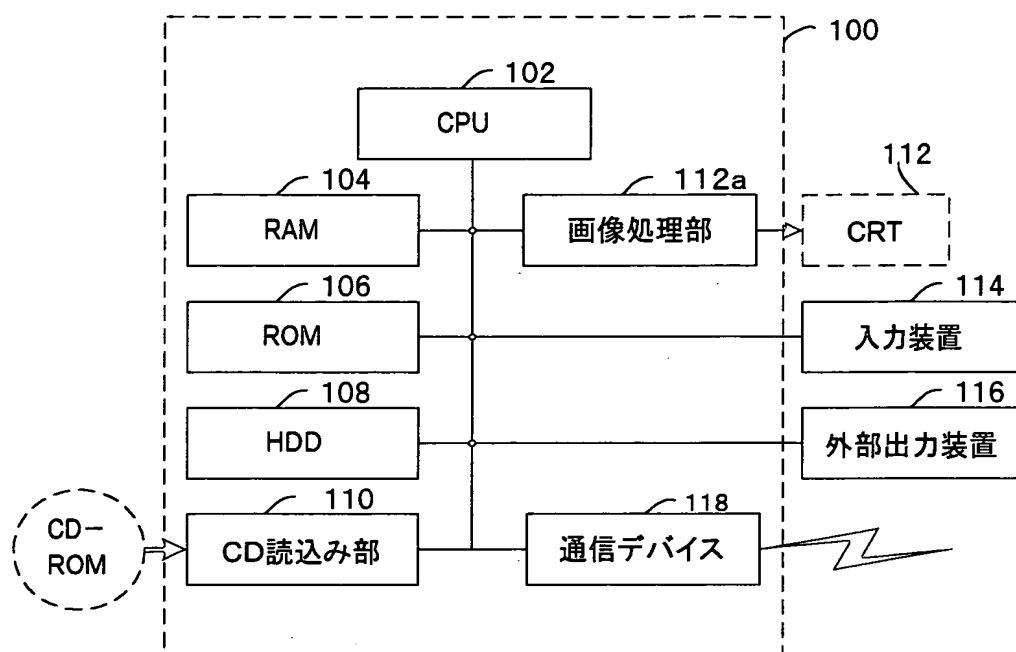
【符号の説明】

1、3 1	原文入力部
2、3 2	形態素解析部
3	変換部
4	定型文検証部
5	置換部
6	語彙情報ファイル
7	パターンマッチングテーブル
8	置き換えテーブル
3 3	構文解析部
3 4	意味解析部
3 5	概念構造変換部
1 0 0	本システム
1 0 2	C P U
1 0 4	R A M
1 0 6	R O M
1 0 8	H D D
1 1 0	C D 読込み部
1 1 2	C R T
1 1 2 a	画像処理部
1 1 4	入力装置
1 1 6	その他の外部出力装置
1 1 8	通信デバイス部

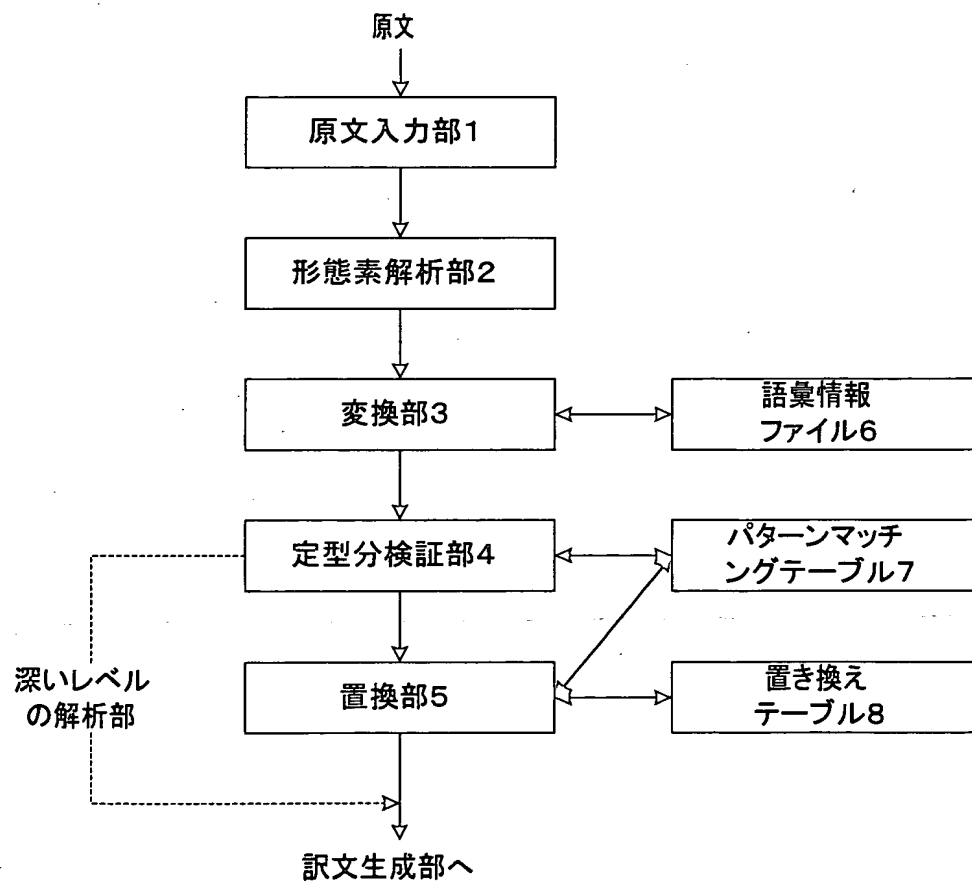
【書類名】

図面

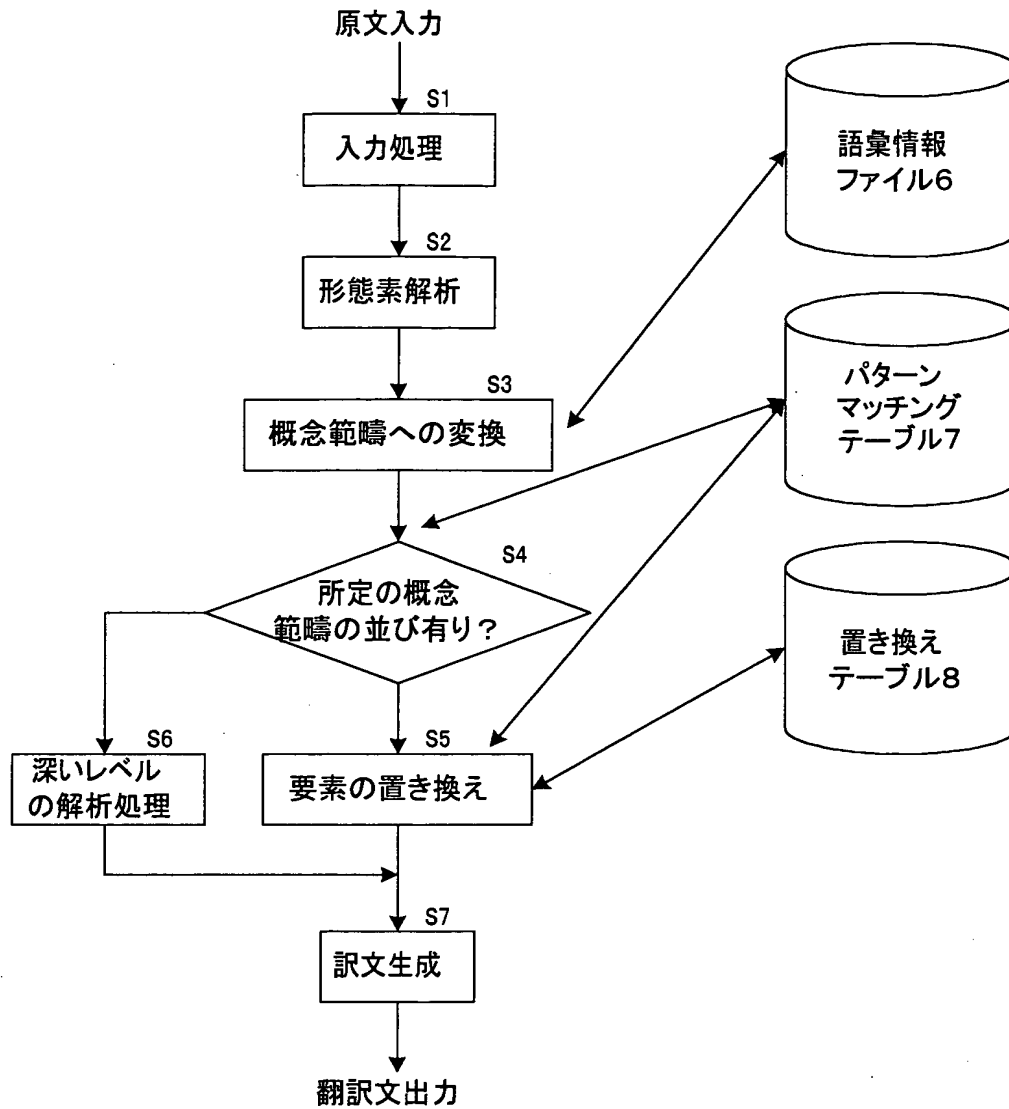
【図 1】



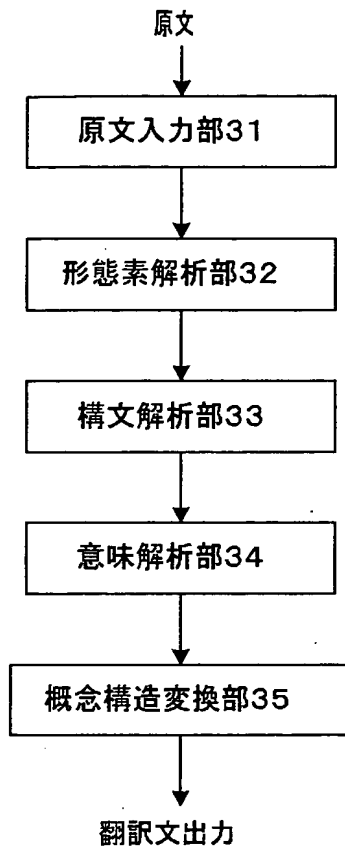
【図 2】



【図 3】



【図 4】



【書類名】 要約書

【要約】

【課題】 定型的な文に対して、パターンマッチング手法により高速に解析処理することができる機械翻訳装置及びそのような翻訳を行うプログラムを記録したコンピュータ読み取り可能な記録媒体を提供せんとするものである。

【解決手段】 原文を入力する原文入力部 1 と、原文入力部 1 に入力された原文を形態素に切り出す形態素解析部 2 と、形態素解析部 2 から切り出された形態素を概念範疇に変換する変換部 3 と、変換部 3 により変換された概念範疇の列中に所定の定型的な概念範疇の並びが存在するか否かを検証する定型文検証部 4 と、上記概念範疇の列中に定型的な概念範疇の並びが存在する時に、概念範疇の並びに対応した所定の訳文の型を生成し、該訳文の型を構成する概念範疇の原文の形態素に対応した訳語に置き換える置換部 5 とを有している。

【選択図】 図 2

出 願 人 履 歴 情 報

識別番号 [000005223]

1. 変更年月日 1996年 3月26日

[変更理由] 住所変更

住 所 神奈川県川崎市中原区上小田中4丁目1番1号

氏 名 富士通株式会社